

The Pennsylvania State University

The Graduate School

College of Education

**AN ANALYSIS OF TEACHER EVALUATION IN CHINA**

A Thesis in

Educational Theory and Policy

By

Yaming Fang

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

May 2019

The thesis of Yaming Fang was reviewed and approved\* by the following:

Mindy L. Kornhaber  
Associate Professor of Education  
Thesis Advisor

David M. Post  
Professor of Education

Kevin Kinser  
Professor of Education  
Head of the Department of Education Policy Studies

\*Signatures are on file in the Graduate School

## **ABSTRACT**

Teacher evaluation in China changed with China's curriculum reform, which focused on nurturing well-rounded students as well as improving the quality of education. In 2009, China implemented its teacher performance pay policy, which is one of the many changes envisioned in the curriculum reform. However, there is no recent research that has been carried out with an aim of establishing whether such teacher evaluation process has been implemented successfully. Building on previous literature on teacher evaluation, this master's thesis used a questionnaire study that examined the teacher evaluation process in public schools from 8 cities in China and the effects of teacher evaluation on teaching practice. The endpoints of this study include explication of the details of the teacher evaluation process in Chinese public schools. In turn, these details will provide the basis for suggesting improvements on the teacher evaluation system and ultimately teacher effectiveness.

## TABLE OF CONTENTS

LIST OF TABLES.....	v
ACKNOWLEDGEMENTS.....	vi
Chapter 1 Introduction .....	1
Chapter 2 Literature Review .....	4
Chapter 3 Data and Methods.....	9
Chapter 4 Results .....	14
Evaluation Methods Based on Insiders of School System.....	14
Data from Students .....	14
Data from Teachers .....	17
Evaluation by School Administrators .....	19
Evaluation by Educational Administrative Department .....	21
Evaluation Methods Based on Outsiders of School System .....	22
Evaluation by Parents.....	22
Evaluation by Education Experts .....	23
Evaluation by Outsourced Companies .....	23
Evaluation by Local Administrative Departments .....	24
Chapter 5 Discussion .....	26
Chapter 6 Conclusion.....	29
Appendix A.....	31
References.....	33

## LIST OF TABLES

<b>Table 3-1:</b> The Number of Teachers or Principals from Different Levels of School in the Sample.....	11
<b>Table 3-2:</b> Brief Descriptions of Participants' School Setting .....	12

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who helped and supported me during the writing of this master's thesis. I gratefully acknowledge the help of my advisor, Dr. Mindy Kornhaber, who has offered me valuable suggestions in my master's study. Without her encouragement, insightful criticism, and gracious guidance, the completion of this thesis would not have been possible.

I owe a special debt of gratitude to Dr. David Post, from whose devoted instruction and enlightening suggestions I have benefited a lot.

I also thank all the other faculties and staff who helped me in my master program. When I encountered any difficulties in my studies and in the completion of this thesis, they helped and supported me without any hesitation.

Finally, I express my gratitude to my beloved parents and friends who have always been there for me, and who have supported me in my life without tiring.

## Chapter 1

### Introduction

Education is the future of a country, and teachers play a vital role in ensuring the prosperity of generations by nurturing the students who pass through their classes. It is important to have an evaluation system that assesses teachers' performance, as teacher performance is an essential component of any education system. Critics of performance payment such as Diane Ravitch strongly oppose this method for improving teachers' performance as it is her belief that, "test score gains will be based on cheating and gaming the system" (2009). However, research has shown that, indeed, performance payment has a positive influence on the performance of teachers. Casson (2007) mentions that teachers with payment linked to their performance put more effort into nurturing students than teachers who work without such incentives. However, without a proper evaluation system, schools cannot tell whether poor performance is caused by students themselves or by insufficient effort of teachers.

Therefore, a teacher evaluation mechanism can be a good way to hold teachers accountable. With a well-designed evaluation system, teachers are more likely to adopt effective teaching methods, which may include cultivating a helpful and friendly classroom atmosphere, proactive lesson management, as well as well-organized lessons with clear objectives (Sammons & Ko, 2008). On the other hand, teachers' perceptions of the evaluation mechanism can also help to explain their job satisfaction (Liu, Stronge, & Xu, 2018).

Under a centralized Ministry of Education, the Chinese national education system is recognized as an efficient system with national standards, a national curriculum, a high-stakes test (college entrance examination), and a clear gateway to transitioning students from one grade to another (Zhao, 2014). The Chinese education system is the largest education system in the world (OECD, 2016). By June 2018, there were 9.75 million students taking the

National College Entrance Examination in China. Investment in education makes up about 4 percent of the total GDP in China. In 1986, the Chinese government passed a compulsory education law, commonly known as nine-year compulsory education, which enables students above six years old to access free education in both primary schools (grades 1 to 6) and middle schools (grades 7 to 9). High school (grades 10 to 12) and college education are not compulsory in China. The Ministry of Education estimates that over 99 percent of school-aged children receive the nine-year compulsory education (China Education Center, 2019).

Generally speaking, the ultimate goal of Chinese high schools is helping students to get high scores on the National College Entrance Examination during the students' graduating year. This test is usually the only criterion for university admission and can only be taken once a year. The criteria for university admissions in China differs from the United States, where university admission decisions are made based on an application package, which consists of the highest SAT or ACT score, a statement of purpose, a resume, letters of recommendation, interview results, athletic or art talent, and TOEFL scores if applicable.

In 1993, the Teachers Law of the People's Republic of China was adopted. It pointed out the expected qualifications for teachers in China,

The State Council shall institute a system of qualifications for teachers. All Chinese citizens who have passed the national teachers' qualification examinations and have educational and teaching ability, may, after being evaluated as qualified, obtain qualifications as teachers. Corresponding records of formal schooling are required.

Teachers' qualification exams consist of two parts: written exams and an interview. In addition, candidates must apply in person and submit identification documents, academic diplomas, a health certificate, a reference letter on the applicant's moral character, and documentation that proves whether he or she has a criminal record (OECD, 2016).

This thesis aims to examine teacher evaluation in China. The two research questions that will guide this thesis are: 1) How do schools evaluate teachers' effectiveness? 2) How do



teachers view the different types of evaluation? The thesis first gives an overview of teacher evaluation through literature reviews. It then provides functions and methods of teacher evaluation and examines different teacher evaluation systems. This thesis will then focus on China's teacher evaluation system by examining the traditional evaluation system as well as teacher evaluation methods after the new curriculum reform was implemented. It uses a questionnaire study that examines the teacher evaluation process in public schools in China and the effects of teacher evaluation on teaching. Principals and teachers from Beijing, Chengdu, Xi'an, Kunming, Xinyang, Xinxiang, Zhumadian, and Shangqiu participated in the questionnaire study. From the questionnaires issued, 225 of them are complete and 13 are incomplete. The incomplete questionnaires have been used in this paper, because they still provide useful information. Studying teacher evaluation in China can provide a way to improve the evaluation processes and ultimately teacher effectiveness and quality. Finally, this paper discusses the benefits and limitations of teacher evaluation methods and finally gives recommendations on how to improve the process.

## Chapter 2

### Literature Review

In general, teacher evaluation has two basic functions. First, it helps teachers focus on improving students' learning performance. Second, it targets improving teacher practice by recognizing strengths and weaknesses for further professional development (Isor é 2009).

Similarly, Teddlie, Stringfield, and Burdett (2003) indicate that teacher evaluation has three major purposes: accountability, promotion and staff development.

In relation to the evaluation methods, Isor é(2009) notes that the following approaches have been used by Economic Co-operation and Development (OECD) countries to evaluate teacher practice and performance: classroom observations, interviews with teachers, teacher-prepared portfolios (video clips, lesson plans, reflection sheets, self-reported questionnaires, samples of student work), student achievement results (absolute performance or value added gains), teacher tests, data from questionnaires and surveys completed by parents and students.

It is worth noting that some large-scale world-wide surveys have also been performed to facilitate the evaluation. For example, in 2009, the Teaching and Learning International Survey (TALIS) within the OECD countries was performed with an aim of assessing teachers' professional competence, related beliefs and attitudes, teachers' classroom practices, professional activities, and teachers' classroom and school level environments.

The traditional teacher evaluation systems relied on a single measure of teacher performance, usually involving perfunctory observation of classroom practices and binary summative ratings. The result was that these factors did not have an impact on teaching practice (Weisberg, Sexton, Mulhern, Keeling, & Schunck, 2009).

Teachers' contributions to student learning should be counted as part of the teacher evaluation system, and the value-added model is often regarded as a tool to achieve this aim (Darling-Hammond, Beardsley, Haertel, & Rothstein, 2012). The value-added model is a

statistical model designed to measure the amount of improvement students make on tests in order to estimate how much growth can be attributed to teachers (Koretz, 2008). The model receives both praise and criticism in the United States. Although the value-added model is generally considered as an objective indicator, it may also suffer from bias in different contexts. For instance, although teachers' performance is an important factor contributing to test scores, they are also influenced by other factors,

Class sizes, curriculum materials, instructional time, availability of resources for learning; home and community supports or challenges; individual student needs and abilities, health, and attendance; peer culture and achievement; prior teachers and schooling, as well as other current teachers; differential summer learning loss; and the specific tests used (Darling-Hammond, et al., 2012).

The National Board for Professional Teaching Standards (NBPTS) designed teaching guide assessments for experienced teachers in the United States. Some states worked together and created the Interstate New Teacher Assessment and Support Consortium (INTASC), which modified these standards for beginner teachers. More than 40 states use these standards for initial teacher licensing (Darling-Hammond, et al., 2012).

Funded by the Bill & Melinda Gates Foundation, the Measures of Effective Teaching (MET) project is based on data from nearly 3,000 volunteer teachers in six school districts in the United States. Its goal was to find out "how evaluation methods could best be used to tell teachers more about the skills that make them most effective and to help districts identify and develop great teaching" (Bill and Melinda Gates Foundation, 2019). The project utilized the following methods to evaluate teachers: student achievement gains on different assessments, classroom observations and teacher reflections, teachers' pedagogical content knowledge, student perceptions of the classroom instructional environment, and teachers' perceptions of working conditions and instructional support at their schools (Bill and Melinda Gates Foundation, 2010).

China has a long history of education. For a long time, under the influence of Confucian tradition, China put great emphasis on the morality of teachers and teachers' professional ethics (Ren, 2009). Teachers should be devoted, patient, professional, diligent, objective, kind-hearted, loving to students, and so on.

Based on Liu and Teddlie's conclusion (2003), China's evaluation system can be divided into two parts: traditional teacher evaluation (the late 1970s to 2001), and contemporary teacher evaluation (2001 to now). According to Ying and Fan (2001), the traditional teacher evaluation system in China is made up of teacher morality, attendance rate, teaching ability, and student performance (which is based on student test scores). However, through the traditional teacher evaluation method, teachers only paid attention to school expectations instead of their own professional development. Also, students' test scores were overemphasized in the traditional teacher evaluation process. The traditional evaluation process also focused too much on quantitative data, though some of the qualities of teaching cannot be adequately represented by quantitative data.

According to Teachers Law of People's Republic of China (1993),

Schools or other institutions of education shall conduct assessment of teachers' political awareness and ideological level, professional competence, attitude toward work and their performances. The administrative departments of education shall guide and supervise the assessment work for teachers. Assessment shall be conducted in an objective, fair and accurate manner and in the process of assessment; opinions from teachers themselves, their colleagues and students shall be taken into full consideration. Assessment results shall be the basis for teachers' appointment and pay rise as well as rewards and punishments.

In 2001, China issued the Curriculum Reform Guidelines for Basic Education. The purpose of the new curriculum reform was to ensure the delivery of well-rounded quality education (State Council, 2008). Since then, although the Ministry of Education still emphasizes the importance of teacher professional ethics, it is also aimed at improving the evaluation of teaching by setting up scientific and diverse benchmarks for such evaluation. New teaching

methods and more active-learner centered approaches have been recommended to the school systems (Jiang & Yao, 2012). Based on Wang and Zhao's conclusion (2013), after the reform, teacher evaluation should make some modifications, including the purposes of the evaluation, the criteria or requirements for the evaluations, the evaluators themselves, and the evaluation measures and methods. The new methods also pointed out the importance of teacher self-evaluation, while also stating that evaluation from principals, peer teachers, students and parents were also necessary.

In 2008, the State Council of China issued "Guidelines to the Implementation of Performance Pay in Compulsory Schools" which rewards high performing teachers with merit pay (State Council, 2008). In order to implement the guidelines, the Chinese Ministry of Education enacted the "Guidelines to Teacher Performance Evaluation in Compulsory Schools" (OECD, 2009). Since then, Chinese schools have developed their own teacher evaluation methods based on these government guidelines. In 2009, a national-wide performance pay evaluation was implemented in China. According to a document published by the Chinese Ministry of Education in 2010, the evaluation focused on developing a moral, ability, and performance-oriented system to evaluate and discover talented educators. Such efforts are not limited to k-12 schools. Colleges, other professional institutions, as well as intermediary agencies are also encouraged to evaluate the level and quality of college disciplines, faculties and courses.

In a meta-analysis conducted by Pham, Nguyen, and Springer (2017), a merit pay program is associated with a statistically significant slightly positive effect on student test scores based on 44 primary studies in the United States. However, in South Korea, the performance pay system was abolished since teachers generally have high salaries, and they know that their jobs will not be replaced by others, thus making the merit incentive unnecessary (Yoo, 2018).

According to the Teachers Law of the People's Republic of China (1993), the State Council needs to institute a system for managing grades of professional titles for teachers. Teachers in China are ranked by professional grades. Similarly, in universities, teachers are ranked by professional grades such as professor, associate professor, assistant professor and lecturer. In primary, middle, and high school there are senior teachers, first-level teachers, second-level teachers, and third-level teachers. Teachers in primary, middle, and high schools use the same title system. Professional titles are beneficial for professional development, and they are symbols of social status (OECD, 2016).

On February 23, 2019, "China's Education Modernization 2035" was issued by the Central Committee of the Communist Party of China and the State Council of China. The guidelines aim to build a team of highly qualified professional and innovative teachers. In order to strengthen teachers' ethics, "China's Education Modernization 2035" regarded teacher ethics as the first criterion for evaluating the quality of teachers. In addition, promoting long-term and systemic teacher ethics is of great importance. Furthermore, attention should be paid to improving the teacher qualification system, the licensure system, as well as teacher professional title, position and evaluation system.

However, there is little research that seeks to understand how the teacher evaluation systems are currently implemented in schools and the impact of these evaluations on teachers' teaching practice. This study tries to address these questions by using a questionnaire that examines teacher evaluation in public schools in China. The first-hand responses from principals and teachers are useful and insightful sources that enable this paper to take a deeper look at real school contexts.

## Chapter 3

### Data and Methods

A questionnaire was developed to examine the teacher evaluation process in public schools in China. The schools covered in this study include elementary schools, middle schools and high schools located in China. Detailed information about how the questionnaires would be involved in this study, the purpose of the study, and the potential risks of being a participant was provided to the participants. All the participants were asked to fill out a questionnaire with the following questions: (1) Is your school an elementary, middle, or high school? (2) What is your role in the school? (3) What is your school size? (4) What is the class size? (5) How does your school evaluate teacher effectiveness? and (6) In what ways does teacher practice change after teacher evaluation? Except for questions (1) and (2), the other questions are open-ended questions. The full version of the questionnaire is attached as Appendix A.

The proposed study was reviewed by The Office for Research Protections, Pennsylvania State University (Study ID STUDY00011602). It was determined that the proposed activity does not require formal IRB review because the research met the criteria for exempt research according to the policies of the institution and the provisions of applicable federal regulations.

This study first used a convenience sample. The author sent questionnaires to her previous teachers and her family members, who used their social networks to send the questionnaire to their friends or relatives who are teachers or principals from Shihe district, Xinyang city, the author's hometown. Some of the teachers and principals found it to be an interesting topic worthy of involvement, and they sent the questionnaire to their colleagues. Eventually, the convenience sample turned into a snowball sample.

From February 21, 2019 to March 21, 2019, the questionnaires were collected. The

author and her family members were the first to send out questionnaires, totaling 20 in number. From these 20, the questionnaire snowballed. The questionnaires were written in Chinese on a Word document and were sent via email, also written in Chinese. From the 250 questionnaires sent back to the author, 225 questionnaires have complete responses. A “complete response” means that all the questions are answered according to the instructions on the questionnaires. Answers such as “I have no idea”, “No”, or “Good” are too short and uninformative, so they are not counted as complete responses. The last question in the questionnaire (“Do you have any suggestions or opinions about teacher evaluation? If any, please describe it in detail.”) was not answered in 13 questionnaires. However, these 13 questionnaires are still used for investigating the previous questions, since respondents still provided complete responses to those previous questions. Usually, complete responses are one or two pages long. The longest response is five pages long.

Of the 105 schools involved in this study, 54 were primary schools, 32 were middle schools, and 19 were high schools. There are some teachers and principals who work together in one school, so the number of schools is less than the number of participants. Responses from 74 principals and 164 teachers are used in this study.

The schools that are the target of this study usually have a class population of 45 to 65 students in urban areas, and 15 to 45 students in suburban or rural areas. One urban primary school in Shangqiu had 9,852 students in total, making it the school with the most students in this study. Conversely, one rural primary school in Zhumadian had only 8 students in total, making it the least populated school in this study.

Participants of this study worked in eight Chinese cities: Beijing, Chengdu, Xi’an, Kunming, Xinyang, Xinxiang, Zhumadian, and Shangqiu. Except for Beijing, every city in this study has at least one school from urban districts and one from suburban or rural districts. Questionnaires from Beijing are only limited to urban districts. Below is a summary of the data sample.



**Table 3-1**  
The Number of Teachers or Principals from Different Levels of School in the Sample

City/Staff	Elementary School	Middle School	High School	Total
Beijing				8
Teachers	2	1	2	5
Principals	1	1	1	3
Chengdu				8
Teachers	3	1	2	6
Principals	0	1	1	2
Kunming				6
Teachers	1	2	1	4
Principals	1	0	1	2
Shangqiu				50
Teachers	19	8	6	33
Principals	7	6	4	17
Xi'an				9
Teachers	2	2	1	7
Principals	1	0	1	2
Xinxiang				52
Teachers	20	14	6	40
Principals	5	4	3	12
Xinyang				48
Teachers	16	12	7	35
Principals	6	3	4	13
Zhumadian				59
Teachers	22	8	6	36
Principals	10	7	6	23

Note: Total number of sample is 238. Total number of teachers in sample is 164. Total number of principals in sample is 74.

**Table 3-2**  
Brief Descriptions of Participants' School Setting

City	Province	Population (in Million)	Number of Elementary Schools	Number of Elementary School Students	Number of Middle Schools	Number of Middle School Students	Number of High Schools	Number of High School Students
Beijing	NA	21	984	876,000	462	266,000	304	164,000
Chengdu	Sichuan	16	569	942,000	458	380,600	143	203,800
Kunming	Yunnan	7	755	497,971	199	220,926	120	116,223
Shangqiu	Henan	7.3	2,115	783,100	405	329,000	34	153,500
Xi'an	Shanxi	9	1,052	666,824	422	259,263	129	158,173
Xinxiang	Henan	6	1,043	632,000	298	273,000	47	107,000
Xinyang	Henan	8.8	1394	674,265	316	332,361	62	187,182
Zhumadian	Henan	8	1946	792,315	303	335,162	35	159,777

Note: Sources credit to Beijing Municipal, 2019, Chengdu Education Bureau, 2019, Kunming Education Bureau, 2019, Admission Office of Shangqiu, 2019, Xi'an Education Bureau, 2019, Xinxiang Education Bureau, 2019, Xinyang Education Bureau, 2019, Zhumadian Education Bureau, 2019.

I analyzed the questionnaires by using comment boxes. The boxes sorted and highlighted data by city, level of school, principal or teacher, methods of evaluation, and positive or negative influences of teacher evaluation on teaching practice.

For methods of evaluation, there are 21 methods in total. They can be divided into two parts depending on whether the evaluation is produced by insiders within the school system or outsiders of the school system. Insiders within the school system include those who work or

study in the school system, including students, teachers, school administrators, and educational administrative department. Outsiders of the school system include parents, experts, outsourced evaluation companies, and local administrative departments. Local administrative departments refer to departments within the local government administration that have an influence on the education system. Such departments include the education department tasked with monitoring the quality of education at the local level.

Participants' strong positive or negative perspective on teacher evaluation are also differentiated in comment boxes. Positive perspectives are defined by words as "like," "love," "beneficial," "good," "excellent," "effective," "useful," and "advantageous." Negative perspectives are defined by words as "dislike," "bad," "disadvantageous," "unfavorable," "unfriendly," "unhelpful," "unprofitable," "inefficient," "waste of time."

Compared to the large population and the huge number of teachers in China, the sample size in this study is small, hence the findings should be considered as illustrative of teachers' views of different evaluation methods rather than definitive. These teachers and principals do not represent their respective cities or their schools. This study only received limited responses. Future research with a larger sample and considering longitudinal effect may be required.

## **Chapter 4**

### **Results**

By gathering open-ended responses from teachers and principals, methods of evaluation were described and teachers' experiences of evaluation were also collected. After analyzing this data based on insiders or outsiders of the school system, the results are as follows:

#### **Evaluation Methods Based on Insiders of School System**

##### **Data from Students**

According to teachers' and principals' responses from the sample, the Student Rating of Teacher Effectiveness is a common way to evaluate teachers. One teacher from a high school in Beijing said,

Students will rate their teacher online anonymously, which includes student satisfaction about the teacher, overall quality of the instruction, classroom management skills, teacher's enthusiasm about teaching, teacher's professional ethics, teacher's professional skills, the workload of assignment, assignment marking, after-class tutorial, the likelihood of recommending this teacher to other students, and so on.

Other teachers' responses about which aspects this method evaluates are similar. Usually, there are four scales in the evaluation: excellent, satisfied, acceptable, and unacceptable. Students were asked to conduct the Student Rating of Teacher Effectiveness at a random time around the end of the semester. One teacher from a middle school in Xi'an said,

The whole class is asked to go to a computer room together and log into their student accounts. In order to protect student privacy, no information about students can be revealed. They answer several multiple-choice questions and open-ended questions related to teachers' teaching performance.

Most of the time, students give their teachers reasonable ratings but chances are, a small number of students click the highest or lowest score buttons regardless of how teachers actually perform, potentially due to students' strong personal preferences. Under these circumstances, most teachers from this study will not focus on those outlier scores. Instead,

they will listen to typical responses and make changes about their teaching to improve teaching practice. One teacher from a high school in Kunming said, “It was not until Student Rating of Teacher Effectiveness, I realized some of my students cannot recognize what I wrote on the blackboard due to my bad handwriting. Then I practiced my writing skills and things got better.”

The second evaluation method based on students is using student examination scores to evaluate teachers. From 94.3 percent of high school teachers’ and principals’ responses, schools use student College Entrance Examination scores to evaluate teachers. The rate of students who pass the national or provincial scores of first tier universities, second tier universities, and third tier universities are important indicators. Teachers who have more students who are admitted to first tier universities have a better chance of being evaluated as good teachers. Just like high schools from this study, most middle school teachers’ and principals’ responses stated that High School Entrance Examination scores are used to evaluate teachers; most elementary school teachers’ and principals’ responses stated that Middle School Entrance Examination scores are used. Schools may also use other examinations to evaluate teachers, including students’ total test scores within the same grade, students’ single subject test scores, and yearly changes in students’ test scores.

Such evaluation is performed following monthly exams, midterm exams, final exams, and mock exams. This method is not only used to evaluate teachers within the same school but also used to evaluate teachers in the same school district where the same tests are usually given to students.

Some of the teachers from this study find this method useful because they can have a better understanding about their teaching results and help each other to improve. One teacher from an elementary school in Chengdu said, “I can have a clear view on whether I have already conveyed knowledge to my students.” Another teacher from a middle school in Xi’an

said, “I can learn the difference between myself and other teachers and get help from them. For example, if my class gets a lower than average score on the exam than other classes, other teachers will tell me to teach a specific knowledge point and the kind of exercises students should do.”

However, some teachers think this method is ineffective because their teaching practice cannot be fully represented by student test scores. Furthermore, teachers have great pressures. One teacher from a high school in Zhumadian said,

I tried my best to teach my students and work diligently, but my students usually performed low in exams. Students in my school are assigned according to their testing scores when they enter this school. The highest 10th percentile students are assigned to a separate class, and the lowest 10 percentile students are assigned to a different, separate class. The other students are assigned randomly. Students in my class are from the lowest 10th percentile of all of the students, so comparing student test scores is of no significance to me.

Another teacher from a middle school in Xinyang said, “I am a new teacher. When exams come, I am more nervous than my students, worrying about whether I am unable to teach them well.”

The third method of evaluation based on students is the student assignment-performance-based method which is similar to the examination-score-based method. One Beijing high school teacher said, “The higher the performance on assignments by students, the higher the rating given to a teacher. I need to pay attention to every process of teaching and learning.”

Related-subject evaluation is another evaluation method. It involves using related subjects to assess the performance of teachers. This method is only reported by teachers from Beijing in this study. Teachers from related subjects influence each other to some extent, thus these related-subjects teachers share their teaching experience with each other at times.

A high school principal said,

For example, physics is related to math. Of students do well in math but do poorly in

physics, a school may conclude that teaching performance of physics teachers are not good. However, if students have low scores in both math and physics, it may conclude that there is a larger issue affecting students.

Dropout rates are also used to evaluate teachers. Schools in rural areas in China usually experience high dropout rates. Students are more eager to work in factories to support their poor families than finishing their studies. Teachers from two rural high schools in Xinyang reported that dropout rates are used to evaluate teachers. If a class has a high number of dropouts, then teachers in this class are likely to be rated lower than teachers whose classes have a low number of dropouts. As a result, teachers often tell their students the importance of finishing their studies so that they can increase their chances of getting good jobs with high salaries.

In addition, teachers in my study from Chengdu indicated that teachers who teach the same subject are evaluated by comparing student test scores, using average scores, the excellent rate and the passing rate in this subject. The head teachers are responsible for students' studies, social behavior, safety, and communication in class. Apart from being evaluated as a single subject teacher, teachers are also evaluated by average scores, the excellent rate and the passing rate of the class they are responsible for. Consequently, head teachers spend more time in dealing with student affairs and know students better than single subject teachers do.

### **Data from Teachers**

Some teachers' responses in the questionnaires indicated that the number of years of teaching experience was used to evaluate teachers. This method simply assumes that with an increase in years of teaching experience, teachers will be more qualified. However, one middle school teacher from Zhumadian argued that some experienced teachers were only experienced in teaching information in the textbook, but they had little knowledge about how to operate

multimedia equipment such as projectors and speakers. He recommended that such teachers should keep up with the fast-changing world of technology so as to be able to use devices to make class more lively and visual.

Peer teacher evaluation is a common approach referred to in the questionnaire responses. According to teachers' and principals' descriptions, teachers are admitted into other teachers' classes in order to observe and give feedback to one another about their performance. Then, they exchange ideas about how to improve teaching skills, class management skills, and pedagogical understanding. In addition, at the end of semester, teachers evaluate one other anonymously. Much like the Student Rating of Teacher Effectiveness, teachers fill out evaluations in a computer room with randomly generated identification codes to represent themselves. Their satisfaction about other individual teachers, overall quality of the class, classroom management skills, teacher's enthusiasm about teaching, teacher's professional ethics, teacher's professional skills, and so on are also evaluated. After the evaluation, except for the principals who have access to all results, teachers can only see their own evaluation results. Different from Student Rating of Teacher Effectiveness, teachers may waive their rights to evaluate others if he or she has a strong positive personal preference or strong negative personal preference towards another teacher.

Under peer teacher evaluation, teachers can show their improved teaching ability and creativity to build their self-confidence. They also have a chance to compare themselves with other teachers, identify problems, solve problems, and ultimately improve teaching quality. It is also beneficial for principals in that they have indicators about teacher hiring, promotion and firing. One teacher from an urban elementary school in Xi'an said that,

After a peer teacher evaluation, other teachers helped me to identify solutions to my problems. I learned all of these class management methods after the teacher evaluation, and they worked! At the same time, my peer teachers encouraged me a lot and highlighted the significance of teacher morality. Thanks to teacher evaluation, now I am a qualified teacher.



Teacher self-evaluation is another common evaluation method, according to teachers in the sample. As reported by teachers and principals, teachers evaluate themselves at the end of semester by filling out a form, on which they state their strengths and weaknesses, feelings and experiences, confusions and questions concerning their daily jobs. After that, teachers submit their self-evaluation form online and exchange self-evaluation results and ideas with one other. One elementary school teacher from Xinxiang pointed out that this evaluation helped her to know herself better and helped her develop a clear professional development plan after self-evaluation.

According to questionnaire responses, the professional knowledge examination is another evaluation method that is frequently used. Teachers with teaching experience less than five years are required to take the professional knowledge examination every year or every two years. These less experienced teachers are also asked to answer the questions from last year's important examinations as part of their professional examination questions. Many teachers noted in their responses that this evaluation is waived for more experienced teachers who are familiar with the examination questions. Rather, they are evaluated based on their education theory, syllabus, curriculum instruction, research project discussion, and their opinions about textbooks. One elementary school teacher from Kunming said, "Syllabus design is the part that I am not satisfied with myself this semester. I hope my classes can be more attractive to students so as to attract their attention. These young kids are easily distracted. I will put more effort in this area."

### **Evaluation by School Administrators**

Principal evaluation is a traditional evaluation method. This method involves principals who randomly choose a teacher's class to observe. The principal then takes notes on the teacher and his or her teaching method. Principals will also look at other indicators of teaching performance, such as student performance, other teachers' comments on this teacher,

experts' comments, student rating of teacher effectiveness, outsourced evaluation company rating, and so on to evaluate teachers. If a teacher's rates are relatively low in many evaluations, the principal will make an appointment with this teacher to discuss the situation, give recommendations and encouragement to help the teacher to overcome the difficulties. New teachers are more likely to meet principals in this situation. On the other hand, if a teacher's ratings are relatively high in many evaluations, the principal will also schedule a meeting with this teacher and ask them to share their experiences with other peer teachers. Under this evaluation, teachers' working enthusiasm is elevated. It is worth noting that one principal from Chengdu suggested that principals should be intelligent, impartial, industrious, reliable, considerate, empathetic, conscientious, and easygoing during evaluations.

One principal from Xinyang reported that her elementary school held a democratic vote to choose high performing teachers. During the annual teacher meeting, the teachers vote for high performing teacher candidates. After the school leaders discuss and authenticate the teachers' qualification, the high performing teachers get a financial bonus at the end of the school year. This incentive method motivates teachers to work hard and be honored.

Performance pay method was adopted in 2009 in China. Principals and teachers noted that based on quantitative assessment of workload, attendance, and teaching quality, teachers are rated as excellent, satisfactory, acceptable, and unacceptable in a percentage system. The final performance pay is based on teachers' rating. This method rewards those who work more and punishes those who lack motivation. One teacher from Xinyang high school said, "This method motivates teachers to work hard and improve on their performance. The method is effective, as it rewards good performance which is central in an education system." However, another middle school teacher from Xi'an regarded this method as ineffective because he thought the merit reward amount was too low, and that there was no significant difference between being rewarded and not being rewarded.

One school principal in Shangqiu established a teacher development database online where he consistently and continually collects teachers' working results and keeps track of every teacher's development. The database mainly records and collects relevant materials such as teachers' work plan, work summary, typical teaching design, teaching case analysis, teaching reflection, multi-subject evaluation records, teaching achievements, award-winning certificates, scientific research results, published papers, and so on. Based on this data, teachers' performance can be made clear, and can be easy to compare. Teachers in the elementary school from the sample also found it objective and useful.

Another school principal in Shangqiu paid evaluated teachers' based on their integrity. The principal paid special attention to teachers' integrity by identifying if teachers take bribes from parents. If parents offer a bribe to a teacher, children from that family will get extra care and training from teachers and even inflated test scores. Bribery was reduced in this school after this evaluation method was enforced. Teachers' ethics improved and a healthy schooling environment was cultivated.

### **Evaluation by Educational Administrative Department**

Local education administration evaluation is another kind of evaluation. As indicated in responses, usually it is the local education bureau that assesses teachers and schools. Such a bureau would be equivalent to the State Department of Education in a state in the U.S.

Administrators randomly choose a class to observe. Just like in the principal evaluation method, administrators will listen to a class, take notes on teachers' professional knowledge and skills, as well as their class management skills. After the observation, administrators rate the teachers. The teachers that are observed will represent the whole teacher body in a school; administrators use the teachers' performance as one indicator to rate the quality of a school. In their questionnaire responses, many teachers were unsatisfied with this method because small mistakes may influence the overall rating of school's quality.

Under the professional title system in China, a teacher's salary is closely linked with his or her professional title. A teacher's salary increases as he or she reaches higher levels of the professional hierarchy. One principal explained that every level of professional title corresponds to different levels of academic degrees, years of working, working results, number of high-quality classes, class topic and class activity as well as award-winning certificates attained. Due to the limited numbers of professional titles available, many potentially qualified teachers may not get their titles until the previous title holders retire. As a result, some teachers in this study mentioned losing the motivation to compete for professional titles and also the incentive to work hard.

According to one principal, in order to be conferred a title, some teachers even cheat in the professional title evaluation by bribing those who assess teachers for professional titles or by buying papers from ghostwriters. It is unfair that some unqualified teachers have a professional title, and accordingly, higher salaries, than some teachers who truly work hard, but do not have the title or salary that matches their efforts and skillsets. Many teachers from this study said their motivation to teach was weakened under professional title evaluations.

### **Evaluation Methods Based on Outsiders of School System**

#### **Evaluation by Parents**

In their responses, 56 teachers stated that their schools use parent evaluations to assess teachers. During the parents' meetings every semester, teachers are rated by parents. Schools distribute questionnaires to parents that ask them give their perspectives about each teacher.

Similar to Student Ratings of Teacher Effectiveness, parents can rate teachers as excellent, satisfactory, acceptable, or unacceptable based on their interactions with teacher or what their children have said about the teachers. In their responses, some teachers doubted the effectiveness of this method, stating that many parents are very busy with their

own jobs, so they have limited attention to focus on their children, let alone teachers. Parents may not provide an effective assessment on teachers, and their judgments are largely influenced by their children's perspective on teachers.

### **Evaluation by Education Experts**

Six schools from Beijing in the sample hire experts to evaluate teachers. These experts are usually leading scholars or academic authorities in their subject. For example, an expert performing the evaluation might be the person who is in charge of question setting for the college entrance examination, with rich teaching and research experience. Every semester, schools randomly choose one or two teachers in each subject to be evaluated by the experts. The teacher who teaches the highest grade is more likely to be chosen for such an evaluation. Experts sit in the back of the classroom and attend one lecture and one exercise review class. The experts also talk with other teachers to acquire a general understanding of teaching performance in the school.

After that, the expert rates the teacher's performance and provides recommendations to all of the teachers in this subject as well as the school's principal. One high school teacher in Beijing said, "I think this evaluation is useful. I can learn a lot of inspiring ideas and feasible skills from experts. At the same time, the teacher whose classes are observed by the experts needs to make a lot of preparations for this evaluation. To some extent, it is also a practice."

### **Evaluation by Outsourced Companies**

Teachers from Beijing in the sample reported their respective schools using outsourced companies to assess teacher effectiveness. In contrast to experts or principals, these outsourced evaluation companies rely heavily on data analysis tools to analyze student average score, the high score rate, the low score rate and so on for every student, teacher, and test question following every important exam. Outsourced evaluation companies provide a

report for each teacher at the end of evaluation. They use bar charts, histograms, pie charts, and radar charts to offer a visual analysis. Based on teachers' feedback, schools hire these companies that use big data technology to evaluate teacher quality. A high school teacher from Beijing said: "Utilizing outsourced companies helped me acquire more accurate data about student performance, so I can know what areas of knowledge I should emphasize."

### **Evaluation by Local Administrative Departments**

Many teachers in Xinyang, Xinxiang, Shangqiu, and Zhumadian from the sample reported being subjects of a local administrative department evaluation. Local administrative departments such as the local fire department, the natural resources conservancy department, the financial department, and so on have evaluations of schools that turn into evaluations of teachers. Teachers are required to fill out all kinds of tables and charts in regard to students, classrooms, teachers, and schools. They are also required to complete online training that is irrelevant to teaching but includes all kinds of safety training and tests. They must also write analysis reports. Local administrative departments have the right to evaluate schools in their own areas, but schools do not have a right to refuse these evaluations.

Most of the teachers in their questionnaire responses expressed a negative perception of this evaluation method. One elementary school teacher from Xinyang said, "This evaluation method is irrelevant to teaching and learning, and is just a waste of time, as it took up so much time from my routine duties." Another teacher from a middle school in Shangqiu said, "This useless evaluation method has nothing to do with improving my student test scores, and I don't think it is my responsibility to take all the safety tests."

Based on questionnaire responses, the teachers in this study who are from Beijing reported the greatest number of different evaluation methods. Most of the schools in this study do not evaluate teachers by only one or two indicators. Instead, they combine these methods to arrive at a comprehensive evaluation. Student test-score based evaluation,

performance pay evaluation, and professional title evaluation are the top three most frequently used evaluations in the sample.

Student-assignment based evaluation, related-subject evaluation, expert evaluation, and outsourced company evaluation are used exclusively for teachers in Beijing in the sample.

Democratic vote evaluation, dropout rate evaluation, local education administration evaluation, and other local administrative department evaluation are methods limited to the teachers from Xinyang, Xinxiang, Zhumadian, and Shangqiu in the sample.

Although most teachers in the sample have a positive attitude towards teacher evaluation, some teachers in the responses said they felt great pressure during teacher evaluation, particularly in big cities, like Beijing, where the competition is fierce. Compounding issues, such as relatively low income, high pressure and workload may cause some psychological problems, for example, anxiety, sadness or even depression. One high school teacher from Beijing stated that,

I, together with my colleagues, work pretty hard. I usually go to school as early as 7 am and go home at 10 pm, and even at lunch time I answer questions that students raise. We are highly professional because of our personal skills, years of experience and annual teacher training. Most of the students are satisfied with us and regard us as responsible or good teachers. We are highly responsible for our students, so there is no need to evaluate us. Some evaluations distract our attention and add pressure to our already heavy workload. In fact, we are too busy to deal with these evaluations. I prefer staying with students and answering their questions rather than being evaluated through bureaucratic evaluations.

According to Paufler (2018), principals also showed concerns about the education evaluation system which has a negative impact on teacher and principal morale. Moreover, principals lack autonomy in evaluating teachers and making staffing decisions.

## Chapter 5

### Discussion

Traditionally, teacher evaluation in China was usually based on subjective evaluations. Principal evaluation and peer teacher rating are common evaluation methods which are based on a few classroom visits and observations every semester. In such a setting, personal relationships between teachers can weaken the credibility of evaluation results due to bias.

After the new curriculum reform, most of the teachers' questionnaire responses mentioned that teacher evaluation in their schools is currently based on a more objective approach, specifically, the students' test scores. However, out-of-school variables such as families' socioeconomic status, the highest education level of the parents, parental involvement in student study, number of siblings at home, and neighborhood environment are all factors that may influence a student's test score. Unfortunately, in many schools, students are not randomly distributed in schools or classes, nor randomly assigned to teachers. Instead, the distribution can be influenced by administrators, teachers, and parents' involvement. For example, students with lower performance rates can all be put into the same class, while more gifted students are put into another classroom. Such a non-random distribution method across classrooms creates bias in teacher evaluation. The bias will be beneficial for some teachers but detrimental to other teachers. A decent teacher may be evaluated as a low-performing teacher simply because of the students in his or her class.

If high-stakes tests such as the national college entrance examination are attached to teacher evaluation, they are poor measures of student achievement. According to Campbell (1979), the use of social indicators in social decision-making increases the risk of corrupting the systems that the indicators are meant to improve. In order to get a higher score, students may cheat during exams and thus threaten the validity of tests. Not only will students cheat, teachers may also cheat by diverting the focus of education from actual learning, to merely



encouraging higher scores on examinations. Teachers may merely focus on how to improve student test scores by teaching for purposes of the examination rather than cultivating students' creativity or critical thinking abilities. In addition, if students are in a non-standardized testing condition, test scores may not be valid enough to be used for teacher evaluation purpose. Furthermore, teacher and student relations can also suffer. While trying to achieve high scores, students may experience great pressure from teachers, for example, in terms of extra assignment or exams. Some teachers may even develop a negative impression of those students who have low test scores, all the time undermining the relationship between students and teachers. Overall, although student test scores are used to evaluate teachers' performance, they should not be used as the sole evaluation approach. As teachers become worried and anxious about being evaluated by students' scores, they may artificially raise their student scores.

Another potential limitation of the current evaluation system is that teachers are compared with other teachers in the same school. Thus, teachers' evaluation results can be greatly influenced by the other teachers in their school. For example, an average performing teacher can be superior in a group of low performing teachers, but at the same time be inferior in a group of high performing teachers. In addition, different school systems have different standards to compare teachers' effectiveness. Similarly, an average performing teacher can be superior to a group of teachers in low evaluation standard schools but at the same time be inferior to a group of teachers in high evaluation standard schools. Currently, there is no united benchmark for schools to use. As such, different standards may serve as a catalyst for teacher turnover as teachers may choose to go from one high evaluation standard school to one low evaluation standard school over time.

In some less developed cities, schools may not provide a high enough salary to attract teachers. As a result, schools try their best to maximize teachers' performance pay by giving

all teachers an equally high rating. Moreover, some schools do not have the right to decide whether teachers should be promoted or fired. As such, the evaluation system exists only by name.

Student Rating of Teacher Effectiveness requires further improvement as well. It may not represent teachers' performance, but students' subjective opinions about teacher and class experience.

Beijing uses most of the evaluation methods as it has the most education resources, besides being the best economy in China. Expert evaluation and outsourced company methods are only used by schools in Beijing, since these two evaluations require a lot of funding from schools. Without a solid economic foundation, schools cannot afford these evaluation methods thus making them inapplicable in schools located in less-developed areas.

As a teacher, the most important task is preparing for teaching, teaching in class, and correcting assignments. Some teachers complain that they have a relatively low salary for the work they do. Many of the tasks are not teaching-related, such as ensuring the school environment is clean. However, teachers must complete these non-teaching related tasks as their results will influence the overall rating of a school. If teachers do not do a good job in these evaluations, the ranking of a school will be undermined, and a series of punishment from local administrative departments will follow. This puts great pressure on teachers and increases their workload.

In the questionnaire responses, some principals pointed out that teacher evaluation is formalist and vague, as it cannot appraise those who work hard and punish those who are lazy and lack motivation. There is no standard to calculate how much a teacher works, and thus the principle of appraising those who work more is far from being realistic. As such, principals in the sample argue that teachers cannot be motivated under current evaluation system.

## Chapter 6

### Conclusion

This study used a qualitative approach to analyze teacher evaluation in China. It is aimed at understanding different types of teacher evaluation methods that are used and how teachers perceive the different types of evaluation methods. The study utilized teachers' and principals' responses from public urban, suburban or rural elementary, middle and high schools in Beijing, Chengdu, Xi'an, Kunming, Xinyang, Xinxiang, Zhumadian, and Shangqiu. Following the new curriculum reform in China, different methods are being used by schools to evaluate teachers.

Evaluation methods based on students, teachers, school administrators, and educational administrative departments are based on insiders' evaluations. Evaluation methods based on parents, education experts, outsourced companies and local administrative departments are by outsiders.

As for how to improve teacher evaluation system, Haertel (2013) gives some suggestions. First, evaluation must be based on sound and appropriate student achievement tests. This means that exams and tests should be carefully monitored and supervised to ensure that students do not cheat, thus risking bias in the evaluation stage. Equally, teachers should be monitored during marking and grading of exams to ensure that they do not award marks unfairly so as to influence the evaluation results in their favor. Unless this is done, evaluations based on student performance cannot be said to be credible and effective.

Second, comparisons should be limited to very similar groups of teachers. For example, a teacher teaching mathematics to the ten brightest students cannot be compared to a teacher teaching mathematics to the ten poorest performing students. Such a move will only favor the teachers whose students have the most preferable qualities, while unfairly disfavoring those with poor students by labelling such teachers as poor performers in their work. In order to ensure that an evaluation process is fair, teachers should only be evaluated in groups only if the

groups possess similar qualities and have a fair playing field.

Third, users of teacher evaluation results must be well trained to interpret them, especially in consequential decisions such as promotion, firing and hiring. If the users are not conversant with the meanings of various results, they risk damaging the reputation of the teaching profession by wrongly targeting teachers, as a result of wrong interpretation of evaluation results. At its best, consequential decisions such as promotion, hiring and firing should be made by a group of users who are conversant with the interpretation of such results.

When evaluations are credible and meet the required standards of use and interpretation, teachers will be more likely to embrace the evaluations, as they will view them as not just an additional burden or a waste of time. As a result, evaluation may serve as a means of improving teacher performance.

## **Appendix A**

### **Questionnaire**

Thank you for considering participating in this research study I am conducting for my Master's thesis at Penn State's Department of Education Policy Studies. My study seeks to understand teacher evaluation in China with the hope that this information will help to improve teacher evaluation systems in the country. Your participation, should you choose to participate, is entirely voluntary. Your names and school names will not be identified in any reporting, and identifiable information will be deleted from any reporting. Should you start the questionnaire, you can choose not to finish it. The estimated time to complete the questionnaire is 20 minutes. If you have any questions about this research you can contact me at [yuf62@psu.edu](mailto:yuf62@psu.edu) or my adviser, Dr. Mindy L. Kornhaber at [mlk20@psu.edu](mailto:mlk20@psu.edu).

1. What kind of school do you work in? Elementary or middle or high school?
2. What is your role at the school (for example, teacher or principal)?
3. How many students attend the school?
4. For principals, on average, what is the number of students in a classroom (average class size)?

For teachers, what is the average number of students in the class or classes you teach?

5. How does school evaluate teacher effectiveness? Please describe it in detail.
6. In what ways does teacher practice change after teacher evaluation? Please describe it in

detail.

7. Do you have any suggestions or opinions about teacher evaluation? If any, please describe it in detail.

Thank you very much for your participation! I will provide a summary report of my master thesis results to you after completion.

## References

- Admission Office of Shangqiu. (2019). Retrieved February 9, 2019, from <http://www.sqzsb.com/>
- Beijing Municipal. (2019). *Beijing Municipal Education Commission*. Retrieved February 27, 2019, from Beijing Municipal: <http://jw.beijing.gov.cn/xxgk/>
- Bill and Melinda Gates Foundation. (2010). *Learning about teaching. Initial findings from the measures of Effective Teaching Project*. New York: Bill and Melinda Gates Foundation. Retrieved March 22, 2019, from <https://docs.gatesfoundation.org/documents/preliminary- findings-research-paper.pdf>
- Bill and Melinda Gates Foundation. (2019). *The measures of effective teaching*. Retrieved March 22, 2019, from Bill and Melinda Gates Foundation: <http://k12education.gatesfoundation.org/blog/measures-of-effective-teaching-met->
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67-90.
- Casson, H. Jr. (2007). Reducing Teacher Moral Hazard in the U.S. Elementary and Secondary Educational System through Merit-pay: An Application of the Principal – Agency Theory, *Forum for Social Economics*, Vol. 36, No. 2, pp 87-95.
- Central Committee of the Communist Party of China and the State Council of China. (2019). China’s Education Modernization 2035. Retrieved March, 22, 2019, from [http://www.gov.cn/zhengce/2019-02/23/content\\_5367987.htm](http://www.gov.cn/zhengce/2019-02/23/content_5367987.htm)
- Chinese Ministry of Education. (2001). Curriculum Reform Guidelines for Basic Education. Retrieved March, 8, 2019, from [http://www.moe.edu.cn/jyb\\_sjzl/moe\\_364/moe\\_302/moe\\_368/tnull\\_4404.html](http://www.moe.edu.cn/jyb_sjzl/moe_364/moe_302/moe_368/tnull_4404.html)
- Chinese Ministry of Education (2008). Guidelines to Teacher Performance Evaluation in Compulsory Schools. Retrieved on March, 8, 2019, from

[http://www.moe.gov.cn/jyb\\_xwfb/s6052/moe\\_838/tnull\\_42745.html](http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/tnull_42745.html)

- Chinese Ministry of Education. (2010). Outline of China's National Plan for Medium and Long-term Education Reform and Development (2010–2020). Retrieved on March, 4, 2019, from [http://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/china\\_national\\_long\\_term\\_educational\\_reform\\_development\\_2010-2020\\_eng.pdf](http://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/china_national_long_term_educational_reform_development_2010-2020_eng.pdf)
- Chengdu Education Bureau. (2019). Retrieved February 27, 2019, from Chengdu Education Bureau: <http://www.cd.edu.gov.cn/>
- Darling-Hammond, H. L., Beardsley, A. A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.  
<https://doi.org/10.1177/003172171209300603>
- Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student scores*. Retrieved March 21, 2019, from The National Press Club:  
<https://eric.ed.gov/?id=ED560957>
- Isore, M. (2009). Teacher Evaluation: Current practices in OECD countries and a literature review. *OECD Education Working Papers*, 1-39.  
<http://dx.doi.org/10.1787/223283631428>
- Jiang, P. P., & Yao, Y. X. (2012). From teacher asking questions to student discovering problems— changing students' learning style in the new curriculum. *Education and Teaching Research*, 26(1), 44-48.
- Koretz, D. M. (2008). *Measuring up: What educational testing really tells us*. Massachusetts: Harvard University Press.
- Kumming Education Bureau. (2019). Retrieved February 27, 2019, from Kumming Education Bureau: <http://jyj.km.gov.cn/>
- Liu, S. J., Stronge, J., & Xu, X. X. (2018). The influences of teachers' perceptions of using student achievement data in evaluation and their self-efficacy on job satisfaction:



- Evidence from China. *Asia Pacific Education Review*, 19(4), 493-509.
- Liu, S., & Teddlie, C. (2003). The ongoing development of teacher evaluation and curriculum reform in the People's Republic of China. *Journal of Personnel Evaluation in Education*, 17(3), 243-261.  
<http://dx.doi.org.ezaccess.libraries.psu.edu/10.1007/s11092-005-2982-x>
- National People's Congress Standing Committee. (1993). Teachers Law of the People's Republic of China. [http://www.npc.gov.cn/englishnpc/Law/2007-12/12/content\\_1383815.htm](http://www.npc.gov.cn/englishnpc/Law/2007-12/12/content_1383815.htm) Retrieved on 03/22/2019.
- OECD. (2009). *Creating effective teaching and learning environments*. Retrieved February 9, 2019, from OECD Publishing: <https://www.oecd.org/education/school/43023606.pdf>
- OECD. (2016). *Education in China: A snapshot*. Retrieved March 22, 2019, from <https://www.oecd.org/china/Education-in-China-a-snapshot.pdf>
- Paufler, N. (2018). Declining, morale, diminishing autonomy, and decreasing value: Principal reflections on a high-stakes teacher evaluation system. *International Journal of Education Policy and Leadership*, 13(8).
- Pham, L., Nguyen, T., & Springer, M. (2017, April 3). Teacher merit pay and student test scores: A Meta-Analysis. Retrieved from Vanderbilt University: <https://my.vanderbilt.edu/matthewspringer/working-papers/>
- Ravitch, D. (2009). What's Wrong with Merit Pay? Retrieved on March 27, 2019, from <https://www.hoover.org/research/whats-wrong-merit-pay>
- Ren, W. (2009). Reinforcement on teachers' morality construction—an eternal subject in educational development. *Asian Culture and History*, 1(4), 180-184.
- Sammons, P., & Ko, J. (2008). *Using Systematic Classroom Observation Schedules to Investigate Effective Teaching: Overview of Quantitative Findings*.
- State Council. (2008, December 21). *Guidelines to the implementation of performance pay in Compulsory Schools*. Retrieved March 5, 2019, from <http://www.gov.cn/ldhd/2008->

12/21/content\_1184109.htm.

Teddlie, C., Stringfield, S., & Burdett, J. (2003). International comparisons of the relationships among educational effectiveness, evaluation and improvement variables: An overview. *Journal of Personnel Evaluation in Education, 17*, 5-20.

Wang, W. J., & Zhao, X. B. (2003). Teacher evaluation reform in the implementation of the new curriculum. *Elementary and Secondary by School Administration, 7*, 15-17.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York: New Teacher Project.

Xi'an Education Bureau. (2019). Retrieved February 27, 2019, from

<http://edu.xa.gov.cn/ptl/index.html>

Xinxiang Education Bureau. (2019). Retrieved February 27, 2019, from

<http://www.xxjy.gov.cn/>

Xinyang Education Bureau. (2019). Retrieved February 9, 2019, from

<http://www.xinyangedu.gov.cn/Item/list.asp?id=1469>

Ying, P. C., & Fan, G. R. (2001). Case Study on teacher evaluation patterns--On the limitations of traditional pattern of teacher evaluation and exploration of a new pattern. *Theory and Practice of Education, 21*(3), 22-25.

Yoo, J. (2018). Evaluating the new teacher evaluation system in South Korea: Case studies of successful implementation, adaptation, and transformation of mandated policy. *Policy Futures in Education, 16*(3), 277-290. doi:<https://doi.org/10.1177/1478210317751274>

Zhao, Y. (2014). *Who's afraid of the big bad dragon : Why china has the best (and worst) education system in the world*. California: Jossey-Bass.

Zhumadian Education Bureau. (2019). Retrieved February 9, 2019, from

[http://www.zmdedu.gov.cn/zhengwugongkai/wenjian tongzhi/list\\_53\\_49.html](http://www.zmdedu.gov.cn/zhengwugongkai/wenjian tongzhi/list_53_49.html)